Journal of Engineering Science and Technology Review 8 (1) (2015) 91 - 95
Special Issue on Econophysics

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Conference Article

Application of continuous-time random walk to statistical arbitrage

Sergey Osmekhin* **and Frédéric Délèze**

Department of Finance and Statistics, Hanken School of Economics, P.O. Box 479, FI-00101, Helsinki, Finland

Abstract

An analytical statistical arbitrage strategy is proposed, where the distribution of the spread is modelled as a continuous-time random walk. Optimal boundaries, computed as a function of the mean and variance of the first-passage time of the spread, maximises an objective function. The predictability of the trading strategy is analysed and contrasted for two forms of continuous-time random walk processes. We found that the waiting-time distribution has a significant impact on the prediction of the expected profit for intraday trading

Keywords: optimal trading strategy, high frequency trading, econophysics, continuous-time random walk, non-Markovian modelling

1 Introduction

Pairs trading, or more generally statistical arbitrage, is a widely known relative pricing strategy commonly used by the proprietary trading desks of investment banks and hedge funds. Based on the principle of mean-reversion, the strategy consists of trading the spread of two co-integrated portfolios. A long position in the spread is taken as the trading pairs departs from its long-term equilibrium. The reverse position is taken as the price difference converges back.

The strategy has considerably evolved since its inception in the 1980s. The development of telecommunication and the generalization of electronic exchanges have drastically changed the way markets operate. Nowadays more than 60% of all executions are generated by fully automated trading strategies resulting in a strong increase in both trading value and market efficiency. Nevertheless, Gatev *et al* 2006 [1] and Perrin [2] have shown that the original pairs trading strategy that was hugely profitable twenty years ago is still profitable today when trading daily.

Numerous extensions of the original model have been proposed recently. Elliott *et al* 2005 [3] provides an analytical framework, where the spread is modelled as an Ornstein-Uhlenbeck process and the value of the spread is estimated by Kalman filtering. Triantafyllopoulous and Montana [4] extended Elliott *et al* algorithm with time-varying coefficients. Bertram 2009 [5] follows the same approach and describes a general optimal trading strategy where the spread follows an Itô´s process and the optimal trading boundaries solve the Fokker-Planck equation of the first-passage time of the spread.

The drawback of these approaches is to assume that the log-prices are normally distributed. Stock returns are known to be leptokurtic and the volatility is not constant over time [6]. A second limitation of traditional research in finance is to sample asset prices at regular time intervals and to only model prices as stochastic processes. Indeed, the time between transactions increases the predictability of asset returns at a market microstructure level. A sudden lack of market activity on a very liquid instrument anticipates market jumps while a very intense period of activity with thin volumes does not move prices.

To circumvent these two limitations, we present an optimal trading strategy tailored for high-frequency trading, where the distribution of the spread follows a Continuous-Time Random Walk (CTRW), which fully takes into account the non-Markovian, fat-tailness and non-local characters of time series. At a market microstructure level, transaction prices are not martingales and the random-walk model is no longer considered to be a complete and a valid description of short-term price dynamics. Indeed, short-run security price changes typically exhibit both extreme dispersion and dependence between successive observations [7]. In addition, the first-order autocorrelations of short-run speculative price changes are usually negative [8]. Moreover, the time between two executions varies and needs to be modelled stochastically.

The paper presents the continuous-time random walk approach to the statistical arbitrage trading frame- work for the first time. Using tick-by-tick prices of the liquid stocks ANZ.AX and ANZ.NZ we

* E-mail address: sergey.osmekhin@gmail.com
ISSN: 1791-2377

found that the waiting time distribution has a significant impact on the expected profit prediction. In the present study we show the importance of the survival probability distribution for the high-frequency trading strategies. The following chapter presents the model: the optimal trading strategy is first described and then the two forms of CTRW used to model the dynamics of the spread are exposed. The third chapter presents the empirical results. The tick-by-tick transactions for the trading pairs is described the model assumptions are presented. Finally, the impact and performance of spread modelling on the optimal trading strategy are discussed.

2 Model

2.1 Optimal trading strategy

One of the ~~key considerations~~ in pairs trading consists of finding optimal barrier levels, which determine the entry and exit levels of the strategy and condition the frequency of the trades. The total time of the strategy is the sum of the expected time it takes for the spread to go from the entry level until the exit level and the time to go from the exit level back to the entry level, i.e. $T_{total} = T_{enter} + T_{exit}$. The trading frequency is therefore $\frac{1}{T_{total}}$. We follow the same approach as Bertram [5] and compute the optimal boundary levels as a function of the first-passage time of the spread. The first-passage time of xt is defined as the first time that the logarithm of the spread process xt reaches an upper boundary $b1$ or a lower boundary $b2$:

$$T_{[b_1,b_2]}(x_0) = \inf_{t_0 \leq t} \{t | b_2 < x_t < b_1; x_0 = x(0)\}$$

Assuming that the probability densities of finding a log - spread x at a future time t, denoted as $p(x, t|x_0, t_0)$, satisfy the absorbing initial conditions $p(x_L, t)=0$ and $p(x_U, t)=0$, $x_L<x_U$, the probability that the process x reaches the boundaries is:

$$G(t; x_U, x_L) = 1 - \int_{x_L}^{x_U} p(x, t|x_0, t_0) dx$$

The corresponding density function solves:

$$g(t|x_0, t_0) = \frac{\partial}{\partial t} G(t|x_0, t_0) = \int_{x_U}^{x_L} p(x, t|x_0, t_0) dx$$

Given two barrier levels $b1$ and $b2$, $b1<b2$, the probability density function of the first-passage time $f(t;b_1, b_2)$ is the convolution of the density function g from the lower limit until the upper limit with the density function g from the lower boundary until the upper limit, i.e.

$$\mathrm{f(t; b_1, b_2)} = \mathrm{g}_{[-\infty,b_2]}(\mathrm{t; b_1, t_0}) \otimes \mathrm{g}_{[b_1,\infty]}(\mathrm{t; b_2, t_0}) \quad (1)$$

The expected trading length solves $E[T_{total}] = \int_0^\infty t f(t; \mathrm{b_1, b_2}) dt$ and the expected trade frequency and variance:

$$E\left[\frac{1}{T_{total}}\right] = \int_0^\infty \frac{1}{t} f(t; \mathrm{b_1, b_2}) \mathrm{dt}$$

$$Var\left[\frac{1}{T_{total}}\right] = \int_0^\infty \frac{1}{t^2} f(t; \mathrm{b_1, b_2}) \mathrm{dt} - E\left[\frac{1}{T_{total}}\right]^2$$

A trading strategy is optimal if the boundaries b_1 and b_2 maximise an objective function, which typically is the expected return of a portfolio μp or its Sharpe ratio. With fixed barriers b_1 and b_2 the return per trade is deterministic $b_2 — b_1 — c$ where c is the transaction cost, but the time between trades is stochastic and depends on the first-passage time of x_t.

The expected profit and variance per trade frequency are:

$$\mu_p = (b_2 - b_1 - c) E\left[\frac{1}{T_{total}}\right] = (b_2 - b_1 - c) \int_0^\infty \frac{1}{t} f(t; \mathrm{b_1, b_2}) \mathrm{dt}$$

$$\sigma_p = (b_2 - b_1 - c)^2 Var\left[\frac{1}{T_{total}}\right] = (b_2 - b_1 - c)^2 \int_0^\infty \frac{1}{t^2} f(t; \mathrm{b_1, b_2}) \mathrm{dt} - \mu_p^2$$

2.2 Continuous-time Random Walk

The log spread process x_t is modelled by CTRW introduced by Montroll and Weiss [9]. We follow the same approach as Scalas *et al* [10, 11, 12] and introduce the following notations:

x(t)=logS(t)	:logarithm of the spread price S at time t.
$\tau_i = t_{i+1} - t_i$	:time between two transactions, also called waiting time.
$\xi_i = x(t_{i+1}) - x(t_i)$	:log-return of asset price S at time t, i.e. $\log\frac{t_{i+1}}{t_i}$
$\varphi(\xi,\tau)$	:joint probability density of returns and waiting-time.
$\psi(\tau) = \int_{-\infty}^{\infty} \varphi(\xi, \tau) d\xi$	:probability density function of the waiting time.
$\lambda(\tau) = \int_0^\infty \varphi(\xi, \tau) d\tau$	:probability density function of asset return.
p(x,t)	:probability density function of finding x at time t.
$\tilde{f}(s) = \int_0^\infty e^{-st} f(t) dt$	:Laplace transform of $f(t)$.

Given a transaction at time t_{i-1}, ψ represents the probability density function that a transaction took place at time $t_{i-1}+\tau$. The probability that a transaction was carried out within $\tau \leq t_i - t_{i-1} \leq \tau + d\tau$ is $\psi(\tau)d\tau$. λ represents the transaction probability density function that the log-price jumped from x to x+ξ. The probability that the log-price did not change during a period greater or equal to τ, also called survival probability until time instant t at the initial position $x = 0$, denoted by $\Psi(\tau)$, is $\Psi(\tau)$=1- $\int_0^{\tau}\psi(t)dt = \int_{\tau}^{\infty}\psi(t)dt$

The probability of finding a log-price x at future time t, $p(x, t)$, solves the following master equation:

$$p(x,t) = \delta(\chi)\Psi(t) + \int_0^{\infty} dt'\psi(t-t')\int_{-\infty}^{\infty}\lambda(x-x')p(x',t')dx' \qquad (2)$$

p(x,0)=δ(x)

An alternative form of the master equation (2) was proposed in Mainardi *et al* [11], which is the solution of the Green function or the fundamental solution of Cauchy problem with the initial condition $p(x, 0) = \delta(x)$:

$$\int_0^t \varphi(t-t')\frac{\partial}{\partial t'}p(x.t')dt' = -p(x,t) + \int_{-\infty}^{\infty}\lambda(x-x')p(x',t)dx' \qquad (3)$$

where the kernel $\varphi(t)$ is defined through its Laplace transform:

$$\bar{\varphi}(s) = \frac{s\bar{\psi}(s)}{1-\bar{\psi}(s)}$$

This form of the master equation is clearly non-Markovian as $\varphi(t)$ is defined as a function of the survival probability. As $\bar{\varphi}(s) = 1$ the master equation for the CTRW becomes Markovian:

$$\frac{\partial}{\partial t}p(x,t) = -p(x,t) + \int_{-\infty}^{\infty}\lambda(x-x')p(x',t)dx' \qquad (4)$$

with p(x,0)=δ(x).

3 Empirical results
3.1 Spread

To test the algorithm with tick-by-tick data, we construct a mean-reverting and stationary portfolio composed of the two listings of the Australia and New Zealand Banking Group Limited (ANZ) [13]. ANZ.AX
is traded in AUD on the Australian stock exchange in Sydney. The ANZ Bank New Zealand Limited, ANZ.NZ, is traded in NZD in Wellington on the New Zealand stock exchange. The time zone difference between the two exchanges is two hours. Using the Australian local time, the spread, x, is computed as follows:

$$x_t = \log(ANZ.AX) - \log(ANZ.NZ) + \log(FX_{AUD/NZD}) \qquad (5)$$

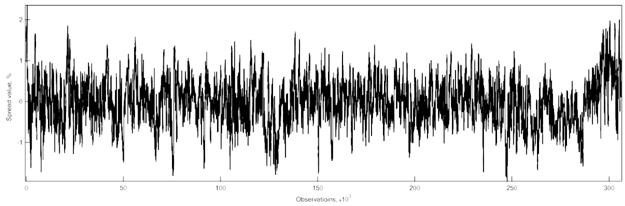
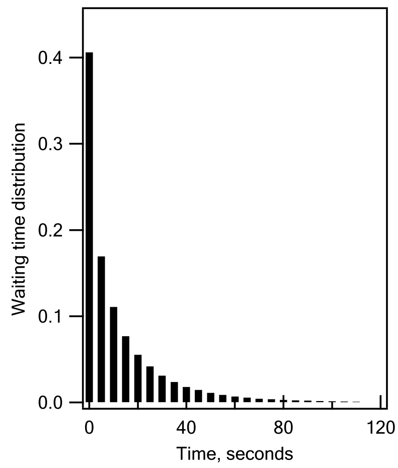
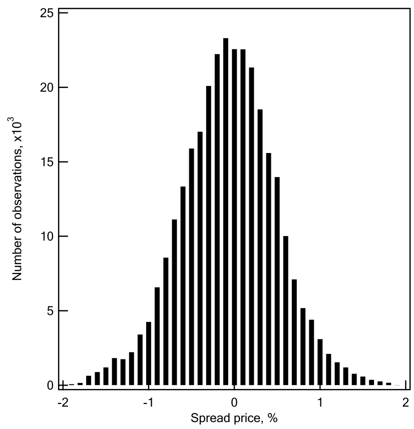


Fig. 1. Spread x_t between ANZ.NZ and ANZ.AX using tick-by-tick trade data covering the period between 4 January 2012 until 8 March 2013.



(a) Waiting-time distribution of x_t



(b) Distribution of the spread x_t

Fig. 2. Waiting-time and asset return distribution of the spread x_t.

The Augmented Dickey-Fuller and the Phillips-Perron tests both reject the null hypothesis that the spread x_t has a unit root at 99% confidence level. Thus, the spread is likely stationary. ANZ.NZ is traded between 10:00 and 16:45 local time and ANZ.AX between 10:00 and 16:00 Sydney time. The spread depicted in figure 1 is computed for the period between 4 January 2012 and 8 March 2013. To prevent information asymmetry, the spread is calculated when both markets are open. It includes 306,071 observations. $\lambda(\xi)$ is typically modelled normally or better with a skewed Student-t distribution to account for the slight excess kurtosis and positive skewness. The time between transactions exhibits a time-decay as illustrated in figure 2a. The intertrade duration at tick-by-tick level is non-exponentially distributed as shown by Scalas *et al* [14] and follows a mixture of compound Poisson processes [15]. The distribution of the spread is close to normal even though the Jarque-Bera test of normality is rejected at 99% confidence level as shown in figure 2b. We do not assume any predefined distribution for the asset return distribution $\lambda(\xi)$ and the waiting-time density function $\psi(\tau)$ but infer them from the tick data.

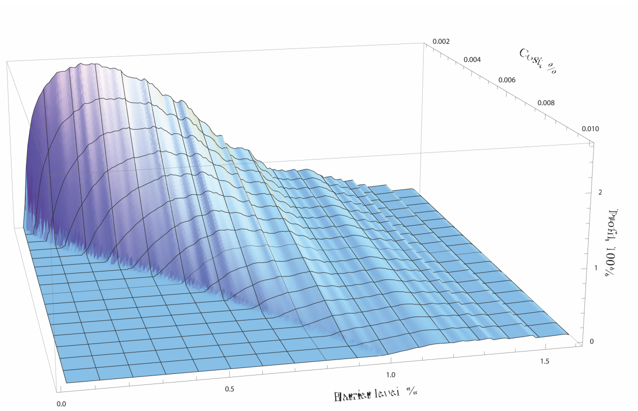


Fig 3. Profitability of the statistical trading strategy as a function of transaction cost and barrier level. When the transaction cost is low, the trade frequency increases but the profit per trade is low.

Table 1. Performance and predictability of the Markovian and non-Markovian approaches comparing to the real profit generated by the strategy.

Cost (%)	Expected return Marko-vian (%)	Barrier level Marko-vian (%)	Expected return non Mark. (%)	Barrier level non-Mark. (%)	Real return (%)	Optimal barrier level (%)	Nb Deals
0.1	1696	0.30	227	0.30	258	0.26	962
0.2	550	0.35	114	0.35	179	0.39	676
0.3	232	0.37	61	0.37	127	0.50	502
0.4	113	0.40	41	0.40	93	0.65	324
0.5	70	0.45	23	0.45	62	0.66	312
0.6	39	0.50	13	0.50	41	0.80	180
0.7	22	0.55	8	0.55	29	0.90	128
0.8	13	0.60	5	0.60	22	1.15	62
0.9	8	0.65	3	0.65	13	1.09	68
1.0	5	0.70	2	0.70	10	1.28	40

3.2 Profitability of the strategy

Figure 3 shows the evolution of the real profit of the trading strategy as a function of the transaction cost and its impact on barrier levels. Stationary spread processes with high speed of mean-reversion built around very liquid financial assets are the best candidates for the described statistical trading strategy. The expected profit is maximal when both transaction cost and barrier levels are low.

Table 1 compares the predictability of the Markovian and non-Markovian approaches of the strategy for a range of transaction costs. The first column represents the transaction cost required to take a long or short position in the spread. The optimal barrier levels for the Markovian (4) and non-Markovian (3) master equations are given in columns 3 and 5. They are identical, meaning that the profit per trade, $b2 — b1 — c$, is the same for both strategies. However, the trade frequency, computed as a function of the first-passage time (1) of the log-price x_t, is model-dependent. It is governed by the probability density function of the asset return $\lambda(\xi)$ and also depends on the waiting-time distribution $\psi(\tau)$ for the non-Markovian approach. The probability of breaching a barrier is higher during high activity than illiquid periods. The impact of the waiting-time distribution on the model precision is particularly important for liquid trading pairs offering a very low transaction cost. When the cost $c = 0.1\%$, the Markovian model is a poor predictor of the real profit generated by the strategy. It forecasts an expected profit of 1696% for the period considered while the actual expected profit reaches 258%. For lower frequency trading the contribution of the waiting-time distribution $\psi(\tau)$ is limited and the non-Markovian equation underestimates the real profit.

4 Conclusion

This paper describes an optimal trading strategy, where the spread is modelled by continuous-time random walk following a non-Markovian process. We found that the waiting-time distribution has a significant impact on the prediction of the expected profit for intraday trading. The non-Markovian approach out- performs the Markovian model for high-frequency trading but underestimates the real profit when trading costs are higher and positions are kept several days. The testing of the statistical trading strategy shows the importance of the survival probability in the optimal trading framework. Non-stationary processes subject to jumps lead to discrepancies between theoretical model predictions and actual observations. Further theoretical work is needed to cover this gap.

References

[1] Evan Gatev, William N. Goetzmann, K. Geert Rouwenhorst, *Pairs Trading: Performance of a Relative-Value Arbitrage Rule*, The Review of Financial Studies 1 v19 n 3 2006

[2] Marcelo Scherer Perrin, *Evaluation of pairs-trading strategy at the Brazilian financial market* Journal of Derivatives and

Hedge Funds 15, 122-136, 2009
[3] Robert J. Elliott, John Van Der Hoek, William P. Malcolm, *Pairs Trading*, Quantitative Finance, 2005 [4] K. Triantafyllopoulos, G. Montana, *Dynamic modeling of mean-reverting spreads for statistical arbi-trage*, Computational Management Science, Volume 8, Numbers 1-2, 2011
[5] William K. Bertram, *Optimal Trading strategies for Itô di↵usion processes* Physica A 388, 2865-2873, 2009
[6] Rosario N. Mantegna and H. Eugene Stanley, *Scaling behaviour in the dynamics of an economic index* Nature vol 376, 1995
[7] Joel Hasbrouck, *Empirical Market Microstructure. The institutions, economics, and econometrics of securities trading* Oxford University Press, 2007
[8] Richard Roll, *A simple implicit measure of the e↵ective bid-ask spread in an eßcient market* Journal of Finance 39, 1127-39, 1984
[9] Elliott W. Montroll and George H. Weiss, *Random Walks on Lattices. II* J. Math. Phys. 6 167, 1965 [10] E. Scalas, R. Gorenflo, F. Mainardi, *Fractional calculus and continuous time finance* Physica A 284 376, 2000
[11] Fransceso Mainardi, Marco Raberto, Rudolf Gorenflo, Enrico Scalas, *Fractional calculus and continuous-time finance II: the waiting-time distribution* Physica A, 287 3-4, 468–481, 2000
[12] Rudolf Gorenflo, Francesco Mainardi, Enrico Scalas and Marco Raberto, *Fractional calculus and continuous-time finance III: the Di↵usion Limit* Mathematical finance, 1710180, Trends Math, Birkhäuser, Basel 2001
[13] Tick-by-tick quote data is provided by tickmarketdata.com
[14] Enrico Scalas, Rudolf Gorenflo, Hugh Luckock, Francesco Mainardi, Maurizio Mantelli, Marco Raberto, *On the Intertrade Waiting-time Distribution* Finance Letters, 2003, **3 (1)**,38-43
[15] Enrico Scalas, *Mixtures of compound Poisson processes as models of tick-by-tick financial data* Chaos, Solitons and Fractals 34 33-40, 2007